

## Programma python

Scrivere una funzione `computeInteractionOverlap(filename, threshold)` che:

- prenda in ingresso un nome di file, con un elenco di interazioni elemento regolatore (RBP o microRNA) - mRNA, e una soglia
- selezioni gli elementi regolatori con almeno `threshold` interazioni, e stampi il numero di interazioni in comune tra ciascuna coppia di elementi regolatori

## File interactions.txt

```
elementName  name  evidenceType
ACO1  hg19_uc001dte.3_3UTR  RIP-chip
ACO1  hg19_uc001dtf.2_3UTR  RIP-chip
ACO1  hg19_uc001dtg.3_3UTR  RIP-chip
ACO1  hg19_uc001nsu.2_3UTR  RIP-chip
...
AGO1  hg19_uc001aaz.2_3UTR  CLIP-seq
AGO1  hg19_uc001abc.2_3UTR  CLIP-seq
AGO1  hg19_uc001abs.2_3UTR  CLIP-se
...
```

## Esempio esecuzione

```
>>> import utility
>>> utility.computeInteractionOverlap('interactions.txt',20000)
reg AGO2 AGO1 ELAVL1 TIAL1 TIA1
AGO2 21711 17867 14358 15654 13133
AGO1 17867 34616 17720 20514 16785
ELAVL1 14358 17720 27045 16806 13940
TIAL1 15654 20514 16806 27091 18368
TIA1 13133 16785 13940 18368 20459
```

## Programma python: suggerimento

Si possono implementare 5 funzioni separate:

1. una che legga il file e restituisca un dizionario con il nome di elemento regolatore come chiave e come valori l'elenco degli mRNA legati
2. una che dato un dizionario di cui sopra e una soglia, ne restituisca uno con solo gli elementi regolatori con piu' interazioni della soglia.
3. una che dato in dizionario di cui sopra, calcoli per ogni coppia di chiavi il numero di elementi in comune tra i loro corrispondenti elenchi (usando una funzione ausiliaria), e ne stampi il valore
4. una ausiliaria (vedi sopra) che dati due elenchi, restituisca il numero di elementi in comune
5. una che realizzi il programma richiesto usando le funzioni di cui sopra

### **Esercizio Shell 1**

Dato il file `interactions.txt`, calcolare quanti *trascritti* sono stati verificati legare una RBP con tecniche di PAR-CLIP

#### *Nota*

Le interazioni riportate sono con UTR (e.g. `hg19_uc010oui.1_3UTR`), dove l'ultimo pezzo (3UTR) indica se si tratta del 3' o del 5'. Lo stesso trascritto puo' essere presente con piu' di un UTR.

#### *Soluzione*

35957

### **Esercizio Shell 2**

Date le sequenze nella directory `fasta`, contare quante proteine includono il seguente motivo:

- un acido aspartico (E) o glutammico (D), seguito da una coppia di cisteine (C) o una coppia di istidine (H), due aminoacidi qualunque e un aminoacido aromatico (F, W o Y).

#### *Soluzione*

3